Stochastic Processes and their Applications 36 (1990) 339-351
North-Holland

NONPARAMETRIC REGRESSION WITH LONG-RANGE DEPENDENCE

Peter HALL

Department of Statistics, Australian National University, Canberra, A.C.T. 2601, Australia

Jeffrey D. HART*

Department of Statistics, Texas A&M University, College Station, TX 77843, USA

Received 4 August 1989

The effect of dependent errors in fixed-design, nonparametric regression is investigated. It is shown that convergence rates for a regression mean estimator under the assumption of independent errors are maintained in the presence of stationary dependent errors, if and only if $\sum r(j) < \infty$, where r is the covariance function. Convergence rates when $\sum r(j) = \infty$ are also investigated. In particular, when the sample is of size n, when the mean function has k derivatives and $r(j) \sim C|j|^{-\alpha}$, the rate is $n^{-k\alpha/(2k+\alpha)}$ for $0 < \alpha < 1$ and $(n^{-1} \log n)^{k/(2k+1)}$ for $\alpha = 1$. These results refer to optimal convergence rates. It is shown that the optimal rates are achieved by kernel estimators.

autoregression * convergence rate * long-range dependence * moving average * nonparametric regression

1. Introduction

The topic of nonparametric regression with dependent errors is being given increasing attention, with most emphasis on circumstances where convergence rates are identical to those in the case of independent errors (Bierens, 1983; Collomb and Härdle, 1986; Hart, 1987, 1990; Truong and Stone, 1988). Here we direct attention at the important complementary problem, in which the dependence of errors is so long-range that classical convergence rates are no longer, or only barely, applicable. We provide a condition on the covariance function which is necessary and sufficient for convergence rates from the independence case to be preserved, and we describe how convergence rates deteriorate when that condition is violated.

Our regression model is

$$Y_i = f(i/n) + \varepsilon_i, \quad 1 \leqslant i \leqslant n,$$

where f is an unknown smooth function and the stochastic process $\{\varepsilon_i, -\infty < i < \infty\}$ is second-order stationary with zero mean; that is, $E(\varepsilon_i) = 0$ and $E(\varepsilon_i \varepsilon_j) = r(i-j)$ for each i and j, where r is the covariance function. This is a reasonable model, for

* Dr Hart's research was done while he was visiting the Australian National University.

0304-4149/90/$03.50

example when the observed data form a time series which is nonstationary in the mean. The classical case is that where $r(j)=0$ for $j\neq 0$, which is usually taken to mean independence. In the absence of independence, attention focuses on the rate at which $r(j)\to 0$ as $|j|\to\infty$. We show that if $r(j)$ is ultimately of the one sign, then the condition $\sum r(j)<\infty$ is necessary and sufficient for optimal estimators of f to have the same convergence rate as in the case of independent errors.

This result is reminiscent of related work on estimation of the mean of a stationary time series, where it is known that the condition $\sum r(j)<\infty$ is necessary and sufficient for $\sqrt{n}$ consistency (Grenander and Szegö, 1958, pp. 206–209). However, in the infinite parameter problem of curve estimation, the convergence rate is always slower than $n^{-1/2}$, even under the assumption of independent errors. For example, the optimal convergence rate with independent errors is $n^{-k/(2k+1)}$, in the case where the mean function f under estimation has k bounded derivatives. For the sake of simplicity we direct our attention to the case $k=2$, although results for the general case are outlined in Remark (2) following Theorem 2.2.

It is of interest to know exactly how convergence rates deteriorate when $\sum r(j)=\infty$. We show that if $r(j)\sim C|j|^{-\alpha}$, where $C\neq 0$ and $0<\alpha<1$, then in the case of estimating a twice-differentiable f the convergence rate $n^{-2/5}$ reduces to $n^{-2\alpha/(4+\alpha)}$. When $r(j)\sim C|j|^{-1}$, the reduced convergence rate is $(n^{-1}\log n)^{2/5}$. These rates are the best possible, and are achieved by kernel estimators. Our results about optimal convergence rates are of minimax type, first used in the context of curve estimation by Farrell (1967, 1972).

It is worth emphasizing that dependence affects our regression setting somewhat differently than it does the correlation model considered by, for example, Bierens (1983). In the latter setting one wishes to estimate a mean function $m(x)=E(Y_t|X_t=x)$ on the basis of dependent, but stationary, *bivariate* observations $(X_1, Y_1),\ldots,(X_n, Y_n)$. The effect of dependence in such cases tends to be less profound than in the model to be investigated here. In particular, our results do not yield the correct convergence rates for kernel estimators in Bierens' setting with long-range dependence. To appreciate how rates differ in Bierens' model, see Hall and Hart (1990) who treat long-range dependence in density estimation.

The reader is referred to Cox (1984) for a survey of work on the structure and application of processes with long-range dependence (see also Mandelbrot and Van Ness (1968) for early work on long memory processes).

2. Convergence rates

2.1. Summary

Section 2.2 defines our kernel estimator and describes its basic features. Section 2.3 defines infinite-order Gaussian autoregressive processes and discusses properties such as invertibility. A class of twice-differentiable mean functions f is introduced in Section 2.4. Our main results are stated in Section 2.5. There, Theorem 2.1 makes use of the Gaussian processes introduced in Section 2.3, while Theorem 2.2 allows more general error distributions.

2.2. Kernel estimator

Take K to be a bounded, nonnegative, symmetric, piecewise continuous, compactly supported function with $\int K = 1$. Define

$$K_h(x) = K(x/h)\left\{n^{-1}\sum_i K(i/hn)\right\}^{-1}.$$

Our kernel estimator of f is $\hat{f}$, defined by

$$\hat{f}(i/n) = n^{-1}\sum_j K_h\{(i-j)/n\}Y_j \tag{2.1}$$

at points i/n and by interpolation elsewhere. As we shall show, the distance between $\hat{f}(i/n)$ and $f(i/n)$ is of larger order than $n^{-1/2}$. It follows that if f has at least one bounded derivative, then the interpolation error is negligible relative to sampling error in estimation of $f(i/n)$.

The quantity $h>0$ appearing in (2.1) is called the bandwidth. It is readily checked that if $r(j)\to 0$ as $j\to\infty$, then the conditions $h\to 0$ and $nh\to\infty$ are necessary and sufficient for bias and variance respectively of the kernel estimator to converge to zero. Consistent estimation of f is not necessarily possible if $r(j)$ does not converge to zero, even if f is constant. To appreciate why, suppose $f\equiv\mu$ (a constant) and the ε-sequence is Gaussian with $r(j)\downarrow c>0$ as $j\uparrow\infty$. Construct a Gaussian sequence $\xi_0, \xi_1, \xi_2, \ldots$ such that $\{\xi_i, i\geqslant 1\}$ is stationary and independent of ξ_0, $E(\xi_i\xi_j)= r(i-j)-c$ and $E(\xi_0^2)=c$. Then $\{\varepsilon_i, i\geqslant 1\}$ has the same distribution as $\{\xi_0+\xi_i, i\geqslant 1\}$. In the model $Y_i=(\mu+\xi_0)+\xi_i$, $i\geqslant 1$, we may consistently estimate $\mu+\xi_0$. However, μ is not estimable from a single infinite realization $Y_1, Y_2, \ldots$.

2.3. Structure of error process

Suppose that the errors ε_j follow an infinite order Gaussian autoregression,

$$\sum_i b_i\varepsilon_{i+j}=\xi_j, \quad -\infty<j<\infty, \tag{2.2}$$

where the ξ_j's are independent standard normal variables and the weights b_j, $-\infty<j<\infty$, are absolutely summable and such that

$$\int_0^\pi |b(\theta)|^{-2}\,\mathrm{d}\theta<\infty, \tag{2.3}$$

where $b(\theta)=\sum_j b_j\,\mathrm{e}^{ij\theta}$. Condition (2.3) ensures that $\operatorname{var}(\varepsilon_j)<\infty$ and that the autoregression (2.2) may be inverted to yield an infinite order moving average,

$$\varepsilon_j=\sum_i a_i\xi_{i+j}, \quad -\infty<j<\infty$$

(Priestley, 1981, pp. 144–145). The weights a_j are expressible in terms of the b_j's via

$$a_j = \pi^{-1} \int_0^{\pi} b(\theta)^{-1} \cos(j\theta)\, d\theta.$$

In this notation, the covariance function of the error process is given by

$$r(j) = E(\varepsilon_o \varepsilon_j) = \sum_i a_i a_{i+j}.$$

Likewise we may define

$$s(j) = \sum_i b_i b_{i+j},$$

$$\rho(\theta) = r(0) + 2 \sum_{j=1}^{\infty} r(j) \cos j\theta,$$

$$\sigma(\theta) = 1/\rho(\theta) = s(0) + 2 \sum_{j=1}^{\infty} s(j) \cos j\theta.$$

The function ρ, of course, is the spectrum of the process $\{\varepsilon_j\}$, and $s(j)$ is the so-called inverse autocovariance function (Priestley, 1981, p. 377). Particularly note that

$$1/\rho(0) = \sigma(0) = \sum_j s(j) = \left(\sum_j b_j\right)^2.$$

Therefore, $\sum r(j)$ converges if and only if $\sum b_j \neq 0$. The first part of Theorem 2.1 shows that if the $r(j)$'s are ultimately of one sign, then the condition $\sum r(j) < \infty$ is necessary and sufficient for the optimal convergence rate of a general estimator of f (not just the kernel estimator $\hat{f}$) to be $n^{-2/5}$.

Throughout this paper we assume $\sum r(j) > 0$. That condition is automatically satisfied when ε_j is defined by the autoregression (2.2), with $\sum |b_j| < \infty$.

2.4. Function class

Our results about optimal convergence rates are framed in terms of the class $\mathscr{C}_2(B)$ of all twice-differentiable functions f on $(0, 1)$ which satisfy

$$\sup_{0<x<1} \max_{j=0,1,2} |f^{(j)}(x)| \leq B.$$

2.5. Main theorems

In the case of Theorem 2.1, assume that the errors ε_j are defined by the Gaussian autoregression (2.2), that (2.3) holds, and that $\sum |b_j| < \infty$. For $x_0 \in (0, 1)$, let $\tilde{f}$ be a general estimator of f on $(0, 1)$, and let $\hat{f}$ be the kernel estimator defined at (2.1), using bandwidth h.

Theorem 2.1. (i) *If $\sum r(j)=\infty$, then there exists a sequence $\lambda_n \to \infty$ such that*

$$\liminf_{n\to\infty} \sup_{f\in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0)-f(x_0)|>\lambda_n n^{-2/5}\}>0. \tag{2.4}$$

Conversely, if $\sum r(j)<\infty$, then for $\Delta>0$ sufficiently small,

$$\liminf_{n\to\infty} \sup_{f\in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0)-f(x_0)|>\Delta n^{-2/5}\}>0. \tag{2.5}$$

If $\sum |r(j)|<\infty$ and h (in the construction of $\hat{f}$) is of size $n^{-1/5}$, then for each $\Delta>0$ we have for $\lambda>0$ sufficiently large,

$$\limsup_{n\to\infty} \sup_{f\in \mathscr{C}_2(B)} \sup_{\Delta<x<1-\Delta} P\{|\hat{f}(x)-f(x)|>\lambda n^{-2/5}\}<\Delta. \tag{2.6}$$

(ii) *Let $0<\alpha\leq 1$, fix $C, C'>0$, and in the cases $0<\alpha<1$ and $\alpha=1$ respectively take $b_j=b_{-j}$,*

$$\begin{aligned} &b_j \sim \begin{cases} C|j|^{-(3-\alpha)/2}, \\ C|j|^{-1}(\log|j|)^{-2}, \end{cases} \\ &\delta(n)=\begin{cases} n^{-2\alpha/(4+\alpha)}, \\ (n^{-1}\log n)^{2/5}, \end{cases} \qquad h\sim h_0=\begin{cases} C'n^{-\alpha/(4+\alpha)}, \\ C'(n^{-1}\log n)^{1/5}, \end{cases} \end{aligned} \tag{2.7}$$

the latter for the construction of $\hat{f}$. Then $r(j)\sim C''|j|^{-\alpha}$ as $|j|\to\infty$, where $C''>0$; for $\Delta>0$ sufficiently small,

$$\liminf_{n\to\infty} \sup_{f\in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0)-f(x_0)|>\Delta\delta(n)\}>0; \tag{2.8}$$

and for each $\Delta>0$ we have for $\lambda>0$ sufficiently large,

$$\limsup_{n\to\infty} \sup_{f\in \mathscr{C}_2(B)} \sup_{\Delta<x<1-\Delta} P\{|\hat{f}(x)-f(x)|>\lambda\delta(n)\}<\Delta. \qquad \square \tag{2.9}$$

Remarks. (1) Result (2.4) establishes that if $\sum r(j)=\infty$, then the convergence rate of $\tilde{f}$ to f is necessarily slower than $n^{-2/5}$. Results (2.5) and (2.6) together show that when $\sum r(j)<\infty$, the optimal convergence rate is $n^{-2/5}$, and that this rate is achieved by kernel estimators. The convergence rate in the case $\sum r(j)=\infty$ is elucidated by results (2.8) and (2.9), which together show that when $r(j)\sim C''|j|^{-\alpha}$, the optimal convergence rate is $n^{-2\alpha/(4+\alpha)}$ (for $0<\alpha<1$) or $(n^{-1}\log n)^{2/5}$ (for $\alpha=1$).

(2) Our proof of Theorem 2.1 is based on a two-point discrimination argument, which shows that for any $\Delta'>0$ the zero on the right-hand side of (2.5) and (2.8) may be replaced by $\frac{1}{2}-\Delta'$ if Δ is taken sufficiently small on the left-hand sides. Use of a more sophisticated multi-point discrimination argument allows the zero to be replaced by $1-\Delta'$.

(3) The arguments leading to (2.6) and (2.9) do not require the underlying process $\{\varepsilon_i, -\infty<i<\infty\}$ to be Gaussian. Furthermore, those results follow by Chebyshev's inequality from general bounds on the mean squared error of $|\hat{f}-f|$, given in Theorem 2.2 below.

Define $\kappa = \frac{1}{2}\int y^2 K(y)\,\mathrm{d}y$, and put

$$v(x) = \begin{cases} x^{-\alpha} C \iint |x-y|^{-\alpha} K(x)K(y)\,\mathrm{d}x\,\mathrm{d}y, \\ x^{-1}(\log x) 2C \int K^2, \\ x^{-1}\rho(0)\int K^2, \end{cases}$$

$$\delta(n) = \begin{cases} n^{-2\alpha/(4+\alpha)}, \\ (n^{-1}\log n)^{2/5}, \\ n^{-2/5}, \end{cases} \qquad h_0 = \begin{cases} C'n^{-\alpha/(4+\alpha)}, \\ C'(n^{-1}\log n)^{1/5}, \\ C'n^{-1/5} \end{cases}$$

in the respective cases $r(j) \sim C|j|^{-\alpha}$ and $0<\alpha<1$, $r(j)\sim C|j|^{-1}$, and $\sum |r(j)| < \infty$ and $\rho(0)>0$.

(4) The assumption $b_j = b_{-j}$ in Theorem 2.1(ii) is for convenience only. It may be replaced by $b_j = 0$ for $j<0$ and b_j satisfying (2.7) as $j \to +\infty$; or by $b_j = 0$ for $j>0$ and b_j satisfying (2.7) as $j\to -\infty$.

Theorem 2.2. (i) *Assume either* $\sum |r(j)| < \infty$ *and* $\rho(0) > 0$, *or* $r(j)\sim C|j|^{-\alpha}$ *and* $0<\alpha \leqslant 1$. *Then for each* $f \in \mathscr{C}_2(B)$ *and each* $\Delta > 0$,

$$E\{\hat f(x) - f(x)\}^2 = h^4\kappa^2\{f''(x)\}^2 + v(nh) + \mathrm{o}\{h^4 + v(nh)\} \tag{2.10}$$

uniformly in $\Delta < x < 1-\Delta$. *The minimum of the right-hand side over h is achieved with a bandwidth of size* h_0, *and is of order* $\delta(n)^2$. *This convergence rate is attained uniformly in f and x, in the sense that*

$$\inf_{h>0} \sup_{f\in\mathscr{C}_2(B)} \sup_{\Delta<x<1-\Delta} E\{\hat f(x) - f(x)\}^2 = \mathrm{O}\{\delta(n)^2\}.$$

If we suppose for the covariance function r only that $r \geqslant 0$, *then a necessary and sufficient condition for*

$$\inf_{h>0} E\{\hat f(x_0) - f(x_0)\}^2 = \mathrm{O}(n^{-4/5})$$

is $\sum r(j) < \infty$.

The proof of Theorem 2.2 is given in Section 3.

Remarks. (1) The contribution of size h^4 in (2.10) comes from squared bias, while the contribution $v(nh)$ derives from variance. An accurate rule of thumb is that for a general covariance function, which is ultimately of one sign, the variance contribution is of size

$$(nh)^{-1}\sum_{j=1}^{nh} |r(j)|.$$

In particular,

$$\sup_{f\in\mathscr{C}_2(B)} \sup_{\Delta<x<1-\Delta} E\{\hat f(x)-f(x)\}^2 = \mathrm{O}\left\{h^4 + (nh)^{-1}\sum_{j=1}^{nh}|r(j)|\right\}.$$

This rule may be shown to hold whenever $r(j)$ is regularly varying in j as $j \to \infty$, or more generally whenever there exists a regularly varying function r^* such that $|r/r^*|$ is bounded away from zero and infinity.

(2) Analogues of both theorems are available for higher function classes. For example, if we define $\mathscr{C}_k(B)$ to be the set of all functions f on $(0, 1)$ satisfying

$$\sup_{0<x<1} \max_{j=0,\dots,k} |f^{(j)}(x)| \leq B,$$

where $k \geq 1$, then Theorem 2.1 continues to hold if we replace $\mathscr{C}_2(B)$ by $\mathscr{C}_k(B)$ throughout, replace $n^{-2/5}$ on the left-hand sides of (2.4) and (2.6) by $n^{-k/(2k+1)}$, take h in the construction of $\hat{f}$ for (2.6) to be $n^{-1/(2k+1)}$, and redefine

$$\delta(n) = \begin{cases} n^{-k\alpha/(2k+\alpha)}, \\ (n^{-1} \log n)^{k/(2k+1)}, \end{cases} \qquad h_0 = \begin{cases} C'n^{-\alpha/(2k+\alpha)}, \\ C'(n^{-1} \log n)^{1/(2k+1)}, \end{cases}$$

in the cases $0 < \alpha < 1$ and $\alpha = 1$, respectively. The estimator $\hat{f}$ should be constructed using a discrete kth order kernel.

(3) Fractional differencing ideas (see, e.g., Granger and Joyeux, 1980) may be used to construct processes having explicitly known covariance functions with the properties $r(j) \sim C''|j|^{-\alpha}$ and $b_j \sim C|j|^{-(3-\alpha)/2}$ (for $0 < \alpha < 1$) or $b_j \sim C|j|^{-1}(\log j)^{-2}$ (for $\alpha = 1$). For example, if we take $d = \frac{1}{2}(1-\alpha)$ in the prescription of Granger and Joyeaux (1980), in which $r(j) = \{\Gamma(1-d)/\Gamma(d)\}\Gamma(j+d)/\Gamma(j+1-d)$, we obtain a process with the desired properties.

3. Proofs

Results (2.6) and (2.9) in Theorem 2.1 follow directly from Theorem 2.2. Therefore we shall only derive results (2.4), (2.5) and (2.8), which comprise Theorems 3.1 and 2.2.

Theorem 3.1. (i) *If $\sum r(j) = \infty$, then there exists a sequence $\lambda_n \to \infty$ such that*

$$\liminf_{n\to\infty} \sup_{f \in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0) - f(x_0)| > \lambda_n n^{-2/5}\} > 0.$$

Conversely, if $\sum r(j) < \infty$, then for $\Delta > 0$ sufficiently small,

$$\liminf_{n\to\infty} \sup_{f \in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0) - f(x_0)| > \Delta n^{-2/5}\} > 0.$$

(ii) *Take $b_j \sim C|j|^{-(3-\alpha)/2}$ (in the case $0 < \alpha < 1$) or $b_j \sim C|j|^{-1}(\log j)^{-3/2}$ (in the case $\alpha = 1$). Then $r(j) \sim C''|j|^{-\alpha}$ as $|j| \to \infty$, where $C'' > 0$ and, for $\Delta > 0$ sufficiently small,*

$$\liminf_{n\to\infty} \sup_{f \in \mathscr{C}_2(B)} P\{|\tilde{f}(x_0) - f(x_0)| > \Delta\delta(n)\} > 0.$$

Proof. Recall from Section 2.2 that interpolation error is of smaller order than estimation error. Therefore we may assume without loss of generality that x_0 is of the form i_0/n, for some i_0. It is notationally convenient to take $x_0 = i_0/n = 0$, so we consider the shifted model where, for some $\Delta > 0$,

$$Y_i = f(i/n) + \varepsilon_i, \quad -\Delta < i/n \leqslant 1 - \Delta,$$

and estimate f at the origin. Indeed, we shall assume that we observe much more than this,

$$Y_i = f(i/n) + \varepsilon_i, \quad -\infty < i < \infty.$$

It turns out that the extra information is of negligible benefit in estimating f.

Let $\psi \geqslant 0$ be a thrice-differentiable function on $(-\infty, \infty)$, vanishing outside $(-1, 1)$ and satisfying $\psi(0) > 0$. Put

$$B' = \sup_{-\infty < x < \infty} \max_{i=0,1,2} |\psi^{(i)}(x)|,$$

and choose $a > 0$ so small that $aB' < B$. Write $m = m(n)$ for a positive integer such that $m \to \infty$ and $h = m/n \to 0$ as $n \to \infty$. Put $f_\theta(x) = \theta a h^2 \psi(x/h)$. Then $f_\theta \in \mathscr{C}_2(B)$ for $\theta = 0$ and $\theta = 1$. Therefore,

$$\sup_{f \in \mathscr{C}_2(B)} P_f\{|\tilde{f}(0) - f(0)| \geqslant \eta\} \geqslant \max_{\theta=0,1} P_{f_\theta}\{|\tilde{f}(0) - f_\theta(0)| \geqslant \eta\}. \tag{3.1}$$

Let $\tilde{\theta} = 0$ or 1 minimize $|\tilde{f}(0) - f_{\tilde{\theta}}(0)|$, and take $\eta = \frac{1}{2} a h^2 \psi(0)$. Then $\tilde{\theta} \neq \theta$ implies $|\tilde{f}(0) - f_\theta(0)| \geqslant \eta$, and so

$$\begin{aligned}
&\max_{\theta=0,1} P_{f_\theta}\{|\tilde{f}(0) - f_\theta(0)| \geqslant \eta\} \\
&\quad \geqslant \max_{\theta=0,1} P_{f_\theta}(\tilde{\theta} \neq \theta) \\
&\quad \geqslant \tfrac{1}{2}\{P_{f_0}(\tilde{\theta} = 1) + P_{f_1}(\tilde{\theta} = 0)\} \\
&\quad \geqslant \tfrac{1}{2}\{P_{f_0}(\hat{\theta} = 1) + P_{f_1}(\hat{\theta} = 0)\},
\end{aligned} \tag{3.2}$$

where $\hat{\theta}$ denotes the maximum likelihood estimator (or likelihood ratio discriminator) in the two-parameter problem. The last inequality follows from the Neyman–Pearson lemma.

Suppose we actually observe the infinite sequence

$$Y_i = f_\theta(i/n) + \varepsilon_i, \quad -\infty < i < \infty.$$

A lower bound to the probability of error in this setting is also a lower bound in the more restrictive model of the theorem (owing to the fact that the lower bound in (3.3) below does not depend on $\tilde{f}$). Put $v_i = \psi(i/m)$, $\boldsymbol{v} = (v_i)$, $\boldsymbol{Y} = (Y_i)$, $\boldsymbol{\varepsilon} = (\varepsilon_i)$ (each doubly infinite column vectors) and $\boldsymbol{V} = (r(i-j))$ (a doubly infinite matrix). Neglecting a multiplicative constant which does not depend on θ, the likelihood of θ is

$$L(\theta) = \exp\{-\tfrac{1}{2}(\boldsymbol{Y} - \theta a h^2 \boldsymbol{v})^{\mathrm{T}} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \theta a h^2 \boldsymbol{v})\}.$$

Therefore,

$$\begin{aligned}P_{f_0}(\hat{\theta}=1) &= P\{L(0)/L(1)<1 \mid \theta=0\}\\ &= P(\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{v} > \tfrac{1}{2}ah^2\boldsymbol{v}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{v})\\ &= 1-\Phi\{\tfrac{1}{2}ah^2(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{v})^{1/2}\}.\end{aligned}$$

Now, $\boldsymbol{V}^{-1}=(s(i-j))$, where $s(i)=\sum_j b_j b_{i+j}$, and so

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{v}=\sum_j\left(\sum_i v_i b_{i+j}\right)^2.$$

Hence,

$$P_{f_0}(\hat{\theta}=1)=1-\Phi\left[\tfrac{1}{2}ah^2\left\{\sum_j\left(\sum_i v_i b_{i+j}\right)^2\right\}^{1/2}\right]=P_{f_1}(\hat{\theta}=0).$$

From this result and from (3.1) and (3.2) we may deduce that

$$\begin{aligned}p &\equiv \sup_{f\in\mathscr{C}_2(B)} P_f\{|\tilde{f}(0)-f(0)|>\tfrac{1}{2}ah^2\psi(0)\}\\ &\geqslant 1-\Phi\left[\tfrac{1}{2}ah^2\left\{\sum_j\left(\sum_i v_i b_{i+j}\right)^2\right\}^{1/2}\right].\end{aligned}\tag{3.3}$$

We treat parts (i) and (ii) of Theorem 3.1 separately.

(i) Suppose that $\sum r(j)=\infty$, that is, $\sum b_j=0$. Since ψ vanishes outside $(-1,1)$, $|v_i|\leqslant C_1 I(|i|\leqslant m)$, where $C_1=\sup\psi$. Writing $C_2=\sum|b_i|$, it follows that

$$\begin{aligned}\sum_j\left(\sum_i v_i b_{i+j}\right)^2 &\leqslant C_1\left(\sup_j\left|\sum_i v_i b_{i+j}\right|\right)\sum_i\sum_j|b_{i+j}|I(|i|\leqslant m)\\ &\leqslant C_1C_2(2m+1)\sup_j\left|\sum_i v_i b_{i+j}\right|.\end{aligned}$$

Since $\sum|b_i|<\infty$, $\sum b_i=0$ and $v_i=\psi(i/m)$,

$$\sup_j\left|\sum_i v_i b_{i+j}\right|\to 0$$

as n (and hence m) $\to\infty$. Therefore, $\sum_j(\sum_i v_i b_{i+j})^2=o(m)$; whence it follows from (3.3) that, for a sequence $\delta_n\to 0$,

$$p\geqslant 1-\Phi\{(nh^5)^{1/2}\delta_n\}.\tag{3.4}$$

Thus, there exists a sequence $\lambda_n\to\infty$ such that

$$\sup_{f\in\mathscr{C}_2(B)} P_f\{|\tilde{f}(0)-f(0)|>n^{-2/5}\lambda_n\}\to\tfrac{1}{2}.$$

(Take $h\sim\nu n^{-1/5}$ in (3.4), for ν arbitrarily large.)

Conversely, suppose that $\sum r(j) < \infty$, that is, $\sum b_j \neq 0$. Now,

$$\sum_j |s(j)| \leq \sum_j \sum_i |b_i||b_{i+j}| = \Big(\sum_i |b_i|\Big)^2$$

and

$$\sum_j \Big(\sum_i v_i b_{i+j}\Big)^2 = (2m+1) \sum_{i=-2m}^{2m} s(i) \frac{1}{2m+1} \sum_{j=-m}^{m} \psi(j/m)\psi\{(j+i)/m\}.$$

Using these facts, dominated convergence, and the fact that $\psi \in \mathscr{C}_2(B)$, it follows that

$$\sum_j \Big(\sum_i v_i b_{i+j}\Big)^2 \sim (2m+1) \int_{-1}^{1} \psi^2(u)\,\mathrm{d}u \sum_i s(i).$$

Take $h = n^{-1/5}$ in (3.3), in which case $h^2 m^{1/2} = 1$, and we see from (3.3) that with $\Delta = \frac{1}{2} a\psi(0)$,

$$\liminf_{n\to\infty} \sup_{f \in \mathscr{C}_2(B)} P_f\{|\tilde{f}(0) - f(0)| > \Delta n^{-2/5}\} > 0.$$

(ii) The symbols $C_1, C_2, \ldots$ denote generic nonzero constants. First we treat the case $0 < \alpha < 1$. Put $\beta = \frac{1}{2}(3-\alpha)$, and let $\{b_j\}$ be an even sequence with $b_j \sim C_1 |j|^{-\beta}$ as $|j| \to \infty$ and early b_j's chosen such that $\sum b_j = 0$. Put

$$b(\theta) = \sum b_j\,\mathrm{e}^{\mathrm{i}j\theta} = b_0 + 2 \sum_{j=1}^{\infty} b_j \cos j\theta = -2 \sum_{j=1}^{\infty} b_j (1 - \cos j\theta),$$

and define

$$a_j = (2\pi)^{-1} \int_{-\pi}^{\pi} b(\theta)^{-1}\,\mathrm{e}^{-\mathrm{i}j\theta}\,\mathrm{d}\theta = \pi^{-1} \int_0^{\pi} b(\theta)^{-1} \cos j\theta\,\mathrm{d}\theta.$$

From Zygmund (1959, Chap. V.2), since $b(\theta) \sim C_2 \theta^{\beta-1}$ as $\theta \downarrow 0$, where $\mathrm{sgn}(C_2) = \mathrm{sgn}(C_1)$, we have $a_j \sim C_3 |j|^{\beta-2}$ as $|j| \to \infty$, where $\mathrm{sgn}(C_3) = \mathrm{sgn}(C_2)$. Similarly,

$$r(j) = \sum_i a_i a_{i+j} = \pi^{-1} \int_0^{\pi} b(\theta)^{-2} \cos j\theta\,\mathrm{d}\theta \sim C_4 |j|^{2(\beta-2)+1} = C_4 |j|^{-\alpha},$$

where $C_4 > 0$. Define

$$v(\theta) = \sum_j v_j\,\mathrm{e}^{\mathrm{i}j\theta} = \sum_j \psi(j/m)\,\mathrm{e}^{\mathrm{i}(j/m)m\theta},$$

and choose ψ to be so smooth that for all m and all $|\theta| \leq m\pi$,

$$\Big|\sum_j \psi(j/m)\,\mathrm{e}^{\mathrm{i}(j/m)\theta}\Big| \leq C_5 m(1+|\theta|)^{-2}.$$

Then $|v(\theta)| \leq C_5 m(1+|m\theta|)^{-2}$; whence (using Parseval's identity)

$$\sum_j \left(\sum_k v_k b_{j+k}\right)^2 = \pi^{-1} \int_0^{\pi} v(\theta)^2 b(\theta)^2 \, d\theta$$

$$= (m\pi)^{-1} \int_0^{m\pi} v(\theta/m)^2 b(\theta/m)^2 \, d\theta$$

$$\leq C_6 m \int_0^{\infty} (1+\theta)^{-4} (\theta/m)^{2\beta-2} \, d\theta$$

$$= C_6 m^{\alpha} \int_0^{\infty} (1+\theta)^{-4} \theta^{2\beta-2} \, d\theta. \tag{3.5}$$

Combining (3.3) and (3.5), and taking $m = hn \sim n^{4/(\alpha+4)}$, we deduce part (ii) of the theorem in the case $0 < \alpha < 1$.

Finally we treat the case $\alpha = 1$. Let $\{b_j\}$ be an even sequence with $b_j \sim C_1 |j|^{-1} (\log |j|)^{-3/2}$ as $j \to \infty$ and early b_j's chosen so that $\sum b_j = 0$. As before, put

$$b(\theta) = \sum b_j \, e^{ij\theta}$$

$$= -2 \sum_{j=1}^{\infty} b_j (1 - \cos j\theta) \sim C_2 (\log \theta^{-1})^{-1/2}$$

as $\theta \downarrow 0$, with $\operatorname{sgn}(C_2) = -\operatorname{sgn}(C_1)$. Then,

$$r(j) = \pi^{-1} \int_0^{\pi} b(\theta)^{-2} \cos j\theta \, d\theta \sim C_3 j^{-1},$$

$$\sum_j \left(\sum_i v_i b_{i+j}\right)^2 \leq C_4 m (\log m)^{-1},$$

where $C_3, C_4 > 0$. Substitute the latter inequality into (3.3), and take $m = hn \sim (n^4 \log n)^{1/5}$ to obtain the desired result. □

Proof of Theorem 2.2. We assume that for some $\Delta > 0$ and all large n, $\min(i, n-i) > \Delta n$. Furthermore we take n to be larger than n_0, where $h < \Delta \min(s_1^{-1}, s_2^{-1})$ for all $n > n_0$ and $(-s_1, s_2)$ is the support of K. Bias and variance contributions to mean squared error are treated separately.

(a) *Bias.* Observe that

$$E\{\hat{f}(i/n)\} - f(i/n)$$

$$= n^{-1} \sum_j K_h(j/n) [f\{(i-j)/n\} - f(i/n) + (j/n) f'(i/n)]. \tag{3.6}$$

If $f \in \mathscr{C}_2(B)$, then the absolute value of the right-hand side is dominated by

$$\tfrac{1}{2} B n^{-1} \sum_j |K_h(j/n)| (j/n)^2 \leq C_1 B h^2. \tag{3.7}$$

If f'' is uniformly continuous, then the right-hand side of (3.6) equals

$$\tfrac{1}{2}f''(i/n)n^{-1}\sum_j K_h(j/n)(j/n)^2+\mathrm{o}(h^2)=\kappa h^2 f''(i/n)+\mathrm{o}(h^2) \tag{3.8}$$

uniformly in $\Delta n<i<n-\Delta n$.

(b) *Variance.* Observe that $\operatorname{var}\{\hat f(i/n)\}$ does not depend on i, and

$$\begin{aligned}\operatorname{var}\{\hat f(i/n)\} &= n^{-2}\sum_j\sum_k K_h(j/n)K_h(k/n)r(j-k) \\ &= \sum_{j=0}^{\infty}\bar r(j)n^{-2}\sum_k K_h(k/n)[K_h\{(k-j)/n\} \\ &\quad + K_h\{(k+j)/n\}],\end{aligned} \tag{3.9}$$

where $\bar r(j)=r(j)$ if $j\geqslant 1$ and $\bar r(0)=\frac{1}{2}r(0)$. If $\sum|r(i)|<\infty$ and $\sum\bar r(i)>0$, then

$$\operatorname{var}\{\hat f(i/n)\}\sim(nh)^{-1}\left\{\sum_{j=0}^{\infty}\bar r(j)\right\}\int K^2.$$

When $r(i)\sim C_0 i^{-\alpha}$ for some $0<\alpha<1$, we see from (3.9) that

$$\begin{aligned}\operatorname{var}\{\hat f(i/n)\} &\sim C_0h^{-2}\iint K(x/h)K(y/h)|n(x-y)|^{-\alpha}\,\mathrm{d}x\,\mathrm{d}y \\ &= C_0(nh)^{-\alpha}\iint K(x)K(y)|x-y|^{-\alpha}\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

When $\alpha=1$ we have for each $c>0$,

$$\begin{aligned}\operatorname{var}\{\hat f(i/n)\} &\sim C_0h^{-2}\iint_{|x-y|>c/n} K(x/h)K(y/h)|n(x-y)|^{-1}\,\mathrm{d}x\,\mathrm{d}y \\ &= C_0(nh)^{-1}\int_{-\infty}^{\infty}K(x)\,\mathrm{d}x\int_{|y|>c/nh}K(x+y)|y|^{-1}\,\mathrm{d}y \\ &\sim 2C_0\left(\int K^2\right)(nh)^{-1}\log(nh).\end{aligned}$$

If $r(i)\geqslant 0$, $\sum r(i)=\infty$ and $K(x)\geqslant\varepsilon>0$ on a nonempty open interval $\mathscr{I}$, then we may conclude from (3.9) that

$$\begin{aligned} nh\operatorname{var}\{\hat f(i/n)\} &\geqslant C_2(nh)^{-1}\sum\sum_{j/nh,k/nh\in\mathscr{I}} r(j-k) \\ &\geqslant C_3\sum_{0\leqslant j\leqslant C_4 nh} r(j)\to\infty \end{aligned} \tag{3.10}$$

as $n\to\infty$.

(c) *Completion.* Using (3.7) to bound the bias of $\hat f$ we conclude from steps (a) and (b) that

$$\sup_{f\in\mathscr{C}_2(B)}\ \sup_{\Delta n<i<n-\Delta n} E\{\hat f(i/n)-f(i/n)\}^2\leqslant C_5\{h^4+v_1(nh)\},$$

where $v_1(nh) = (nh)^{-1}$ if $\sum |r(i)| < \infty$, $(nh)^{-1}\log(nh)$ if $r(i) \sim C_0 i^{-1}$, $(nh)^{-\alpha}$ if $r(i) \sim C_0 i^{-\alpha}$ and $0 < \alpha < 1$. The infimum over h of the right-hand side is dominated by $C_6 n^{-4/5}$, $C_6(n^{-1}\log n)^{4/5}$, $C_6 n^{-4\alpha/(4+\alpha)}$ in these respective cases.

Should f'' be continuous, then, using (3.8) to estimate bias, we find that

$$E\{\hat{f}(i/n) - f(i/n)\}^2 = h^4\kappa^2 f''(i/n)^2 + v(nh) + o\{h^4 + v(nh)\}$$

uniformly in $\min(i, n-i) > \Delta n$. That v admits the formulae ascribed to it in the statement of the theorem follows from the estimates in step (b). In the case where $r(i) \geq 0$ and $\sum r(i) = \infty$ we may deduce from (3.8) and (3.10) that $n^{4/5}\inf_h E(\hat{f} - f)^2 \to \infty$. □

References

Bierens, H.J., Uniform consistency of kernel estimators of a regression function under generalized conditions, J. Amer. Statist. Assoc. 78 (1983) 699-707.

Collomb, G. and W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis, Stochastic Process. Appl. 23 (1986) 77-89.

Cox, D.R., Long-range dependence: A review, in: H.A. David and H.T. David, eds., Proc. 50th Anniversary Conf. Iowa State Statistical Laboratory (Iowa State University Press, 1984) pp. 55-74.

Farrell, R.H., On the lack of a uniformly consistent sequence of estimators of a density function in certain cases, Ann. Math. Statist. 38 (1967) 471-474.

Farrell, R.H., On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, Ann. Math. Statist. 43 (1972) 170-180.

Granger, C.W.J. and R. Joyeaux, An introduction to long-memory time series models and fractional differencing, J. Time Series Anal. 1 (1980) 15-29.

Grenander, U. and G. Szegö, Toeplitz Forms and Their Applications (University of California Press, 1958).

Hall, P. and J.D. Hart, Convergence rates in density estimation for data from infinite-order moving average processes, Probab. Theory Rel. Fields (1990) to appear.

Hart, J.D., Kernel smoothing when the observations are correlated, Tech. Rep. No. 35, Dept. of Statistics, Texas A&M Univ. (College Station, TX, 1987).

Hart, J.D., Kernel regression estimation with time series errors, J. Roy. Statist. Soc. Ser. B. (1990, to appear).

Mandelbrot, B.B. and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. 10 (1968) 422-437.

Priestley, M.B., Spectral Analysis and Time Series (Academic Press, New York, 1981).

Truong, Y.K. and C. Stone, Nonparametric time series prediction: Kernel estimates based on local averages, unpublished manuscript (1988).

Zygmund, A., Trigonometric Series, Vol. 1 (Cambridge University Press, 1959).